Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Journal of **Comparative Effectiveness Research**

Using UK Biobank for sexual minority health research

Catherine L Saunders*,[1] & Efthalia Massou[2]

[1]Senior Research Associate, The Primary Care Unit, Department of Public Health & Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge CB2 0SR, UK

[2]Research Associate, The Primary Care Unit, Department of Public Health & Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge CB2 0SR, UK

*Author for correspondence: ks659@medschl.cam.ac.uk

Aim: Despite poorer health and healthcare outcomes experienced by lesbian, gay and bisexual adults, data for research to characterize and address these disparities remain limited. **Patients & methods:** We describe sexual history information from 502,543 UK Biobank participants recruited between 2006 and 2010, as sexual identity was not collected from the cohort at baseline, and compare this with sexual history and sexual identity responses to the third National Survey of Sexual Attitudes and Lifestyles (NATSAL-3), collected between 2010 and 2012. **Results:** After exclusions, 700 (0.3%) women and 2112 (1.2%) men in UK Biobank reported a history of exclusively same-sex sex and 5162 (2.3%) women and 4275 (2.3%) men reported a history of sex with both women and men; estimates were consistent with, although slightly lower those from NATSAL-3. **Conclusion:** UK Biobank is an important resource for sexual minority health research.

Lay abstract: Although lesbian, gay and bisexual adults experience poorer health and healthcare outcomes, data for researchers to use to understand and address these disparities remain limited. UK Biobank is a research study which recruited, and is now following up, over half a million adults aged between 40 and 70. When participants were recruited they were not asked about sexual identity but they were asked about their history of sex with both women and men. In this research, we compared the responses to this question in UK Biobank to responses from another survey, the third National Survey of Sexual Attitudes and Lifestyles, and found the estimates were broadly comparable. We also provide some advice to researchers on how to use UK Biobank in future research to explore disparities in health outcomes experienced by lesbian, gay and bisexual adults.

First draft submitted: 25 February 2021; Accepted for publication: 1 June 2021; Published online: 7 July 2021

Keywords: cohort description • sexual minority • UK Biobank

Lesbian, gay and bisexual women and men experience poorer health, and poorer healthcare outcomes, compared with heterosexual women and men [1]. In addition to sexual health and mental health disparities, long-term health conditions and cancer, are increasingly identified as priority areas for sexual minority health improvement programs [2–4].

In 2011, the Institute of Medicine highlighted the need for research, including basic epidemiological research, to address these health inequalities [5]. In the UK, the 2018 National Government Equalities Office LGBT Action Plan identified the need for research into the health and healthcare outcomes of sexual minority women and men [6]; and there have been further calls from health charities for better data [4]. UK research resources are likely to become increasingly important especially for periods where US sexual orientation data collections were rolled back [7].

Routinely collected health record data remain the key resource for observational health research [8], however in both the UK [9], and the US [10], recording of sexual orientation information remains limited. The introduction of a mandate for equalities monitoring with the UK Equality Act in 2010 [11] improved the collection of sexual orientation information in survey data [12]. However, adequate sample size for research, particularly among minority

Future Medicine

groups, remains challenging [13]. Despite their limitations, nonpopulation-based sampling methods including convenience and purposive sampling strategies remain important methodologies [14,15].

UK Biobank is a unique resource designed to study the genetic and environmental determinants of chronic and long-term health conditions including cardiovascular disease and cancer [16]. Half a million UK Biobank participants were recruited between 2006 and 2010 and data have been linked to electronic patient records and disease registries. More recently, linkage to COVID-19 data have provided a further valuable dimension. Although there is evidence of a 'healthy volunteer' bias among UK Biobank participants [17], women and men from across the UK are included in the cohort.

Lesbian, gay, bisexual or other sexual orientation or identity was not collected as part of the UK Biobank baseline assessment, however, information on sexual behavior was. This information has been used to date in three research studies, one exploring the genetic determinants of having had sex with someone of the same sex [18], the second exploring risk factors for prostate cancer [19] and the third exploring cervical cancer screening [20]. No study to date has focused on the use of UK Biobank as a resource for research designed to characterize and develop approaches to address disparities in health outcomes experienced by sexual minority women and men.

Therefore, in the context of this pressing evidential need for data to address the health disparities experienced by sexual minority women and men, we evaluated UK Biobank as a resource for research for sexual minority health. We specifically consider sample size, and the reliability of sexual history responses. In addition, we compare responses with responses to the third National Survey of Sexual Attitudes and Lifestyles (NATSAL-3), which is one of the largest and most detailed studies of sexual behavior in the world [21], and explore whether the measure of sexual behavior recorded in UK Biobank can be used to understand disparities experienced by sexual identity and sexual orientation.

In this work, we use the term sexual orientation when discussing health disparities in line with UK Office of National Statistics reporting which aligns with the UK Equality Act [22]. Sexual orientation is an umbrella term that encompasses sexual identity, attraction and behavior [22], which are overlapping but distinct constructs [23,24].

Methods

Data

UK Biobank recruited 503,325 participants aged between 40 and 69 years from 2006 to 2010 in 22 assessment centers across the UK. All people registered with a National Health Service family doctor and living within approximately 25 miles of an assessment center were invited. From 9.2 million invitations, the recruitment response rate was 5.5%. Participants provided detailed information at baseline and follow-up assessments, and provided samples and consent for linkage. Further methodological details are available at www.ukbiobank.ac.uk. The data used in this study were accessed through UK Biobank (application number 42861). Sexual history information was collected using a self-completion touch-screen tool.

NATSAL-3 is a nationally representative survey of women and men living in private households in the UK. Data collection was between September 2010 and August 2012 [21]. History of same- and opposite-sex sexual behavior was collected through a self-completion questionnaire. Access to these data was obtained through UK Data Archive (project ID: 176077) [25]; respondents aged 40–69 were included for comparability with the UK Biobank sample.

Sexual behavior & sexual identity

In the UK Biobank baseline assessment, participants were asked a series of questions about their sexual history.

The initial question asked participants whether they were willing to continue to the sexual history questions: ('*The next section contains questions about your sexual history. If you feel that a question is too sensitive, you can skip the questions or skip the entire section if you prefer*').

Respondents were then asked: '*What was your age when you first had sexual intercourse? (Sexual intercourse includes vaginal, oral or anal intercourse)*,' with an additional response option allowing respondents to identify that they had never had sex.

Respondents were additionally asked: '*Have you ever had sexual intercourse with someone of the same sex?*,' '*How many sexual partners of the same-sex have you had in your lifetime?*' and '*About how many sexual partners have you had in your lifetime?*' There were additional responses to this question, 'Do not know' and 'Prefer not to say.'

Based on responses to these questions about numbers of lifetime sexual partners and numbers of same-sex sexual partners, we identified women and men who reported sex exclusively with men and women, respectively (WSEM

and MSEW). We additionally identified women and men who reported exclusively same-sex sex (WSEW and MSEM) and women and men who reported sex with both women and men (WSWM and MSWM).

A subgroup of UK Biobank participants were asked the sexual history questions again at a repeat assessment visit in 2012/3 and a further subgroup were asked again at a further assessment occurring from 2014 onward.

Using NATSAL-3, we considered responses to the question: '*Sexual experience is any kind of contact with another person that you felt was sexual (it could be just kissing or touching, or intercourse or any other form of sex). I have had some sexual experience...* ,' with response options paired for sex with both men and women 'never'/'only,' 'at least once'/'more often' and 'about equally often.' In addition, respondents in NATSAL-3 were asked '*Which of the options on this card best describes how you think of yourself?*' with response options 'Heterosexual/straight,' 'gay/lesbian,' 'bisexual' and 'other.'

Additional measures

In both UK Biobank and NATSAL-3, age and sex at baseline assessment were recorded, and the Townsend deprivation index (a small area measure, which is scaled to a mean of 0 in the population, with lower or negative values indicating less deprived areas) was used as a measure of material deprivation for respondents from UK Biobank. The index is constructed from the following four census variables, households without a car, overcrowded households, households not owner-occupied, persons unemployed, and in the UK it is widely used in identifying areas of the greatest social need in resource allocation decisions [26]. The Townsend deprivation index was calculated immediately prior to participant joining UK Biobank based on the preceding national census output areas; each participant is assigned a score corresponding to the output area in which their postcode is located [27].

Analysis

In our first set of analyses, we calculated the numbers of UK Biobank participants for whom exclusively same-sex, exclusively opposite-sex and both same- and opposite-sex sexual behavior could be described.

Among women and men who reported sexual history at both the baseline UK Biobank assessment and then again at either the first repeat assessment visit or, if these data were not missing, the second, we calculated longitudinal consistency for exclusively same-sex, exclusively opposite-sex and both same- and opposite-sex sexual behavior.

We additionally described the age, deprivation and lifetime number of sexual partners of UK Biobank participants, stratified by sex and sexual history.

In our second set of analyses, we compared reported sexual behavior in UK Biobank with estimates from NATSAL-3, using the survey weights from NATSAL-3 to give nationally representative estimates in the UK population. We used NATSAL-3 to estimate the positive predictive value (PPV), and sensitivity of sexual behavior as a measure of sexual identity. These PPV are the percentage of women and men in each group defined by sexual behavior who identify with the most closely related measured sexual orientation and these measures of sensitivity present the percentage of women and men who identify with each sexual orientation who report the most closely related sexual behavior history.

Finally, we produced a set of weights that will allow researchers to explore health disparities based on sexual behavior using UK Biobank and use these weighted measures to estimate of disparities experienced by gay, lesbian, bisexual and other sexual minority women and men.

Patient & public involvement

The lay abstract for this work was revised and improved following comments from members of an LGBTQ+ research participation panel [28].

Results

Of the 502,543 UK Biobank participants at baseline assessment, 47,018 (9.4%) did not answer any sexual history questions and 4230 (0.9%) reported never having had sex. An additional 2450 (0.5%) respondents did not give an informative response to the question asking about same-sex sexual behavior; 448,845 (89.3%) respondents (242,222 women and 206,623 men) gave an informative response to the question about whether they had ever had same-sex sex, or not. A further 45,117 (9.0%) participants had missing/uninformative responses to questions about lifetime numbers of same-sex and opposite sex partners. At baseline assessment, after excluding respondents with missing data, 403,728 (80.3%) of UK Biobank participants could therefore be grouped into three cohorts based on their sexual behavior history (Table 1).

Table 1. Sample size and exclusions: numbers of responses to sexual history questions in UK Biobank.

Participant flow through the study	Included (n)			Excluded (n)		
	All	Women	Men	All	Women	Men
All UK Biobank participants at baseline assessment	502,543	273,405	229,138			
Excluding nonresponders to the entire sexual history section	455,525	245,509	210,016	47,018	27,896	19,122
Excluding participants who have never had sex	451,295	243,259	208,036	4230	2250	1980
Excluding participants with uninformative responses to the question about whether they had ever had same-sex sex	448,845	242,222	206,623	2450	1037	1413
Excluding participants with missing/uninformative responses to questions about lifetime numbers of sexual partners and/or lifetime numbers of same-sex partners	403,728	220,714	183,014	45,117	21,508	23,609
Final cohort sizes						
– Exclusively opposite-sex†	391,479	214,852	176,627			
– Exclusively same-sex	2812	700	2112			
– Both same- and opposite-sex	9437	5162	4275			

†We acknowledge that in this and subsequent tables the use of the term 'opposite-sex' reinforces a gender binary which in many cases is not helpful. However, in this work, we use the terminology 'same-sex' and 'opposite-sex' reporting responses based on the wording of the questions asked.
n: Number.

Table 2. Longitudinal consistency of reported same-sex/opposite-sex sexual behavior in UK Biobank.

	Reporting sexual behavior at baseline – n	Where reporting is consistent at follow-up assessment – n (%)
Women		
– All with at least two repeated measures of sexual behavior†	17,539	
– Exclusively opposite-sex	16,842	16,755 (99.5)
– Exclusively same-sex	59	38 (64.4)
– Both same- and opposite-sex	463	412 (89.0)
Men		
– All with at least two repeated measures of sexual behavior†	16,807	
– Exclusively opposite-sex	16,105	15,987 (99.3)
– Exclusively same-sex	195	151 (77.4)
– Both same- and opposite-sex	337	248 (73.6)

†This number also includes people who report never having had sex (175 women and 170 men at baseline).
n: Number.

The majority of UK Biobank participants 391,479 (97.0% of those without missing data) report an exclusively opposite-sex sexual history; 700 (0.3%) women (WSEW) and 2112 (1.2%) men (MSEM) in UK Biobank reported a sexual history of exclusively same-sex sex and 5162 (2.3%) women (WSWM) and 4275 (2.3%) men (MSWM) reported a sexual history of sex with both men and women (Table 1). Among the 448,845 participants with an informative response to the question about ever having had same-sex sex, in total 6924 (2.9%) women and 8895 (4.3%) men reported ever having had sex with someone of the same sex.

Longitudinal consistency of sexual behavior in UK Biobank among the 17,539 women and 16,807 men with repeated measures was good (Table 2), with overall agreement between assessments of 99.0% in women and 98.4% in men (Kappa 0.88 in women and 0.81 in men). Agreement was higher among WSEM (99.5%) and MSEW (99.3%) compared with WSEW (64.4%) and MSEM (77.4%), and WSWM (89.0%) and MSWM (73.6%).

Cohort characteristics are presented in Table 3. Both women and men who report exclusively same-sex or both same- and opposite-sex sexual histories have a younger median age (50 and 51 in women and 51 and 54 in men) than those reporting an exclusively opposite-sex sexual history (median age of 57 in women and 58 in men). WSEM and MSEW live in more affluent areas than women or men with any history of same-sex sex. WSEW report a median of three lifetime sexual partners and WSMW a median of ten; MSEM report a median of 15 lifetime

Table 3. Cohort characteristics (UK Biobank).

	Exclusively opposite-sex	Exclusively same-sex	Both same- and opposite-sex	Never had sex	Missing
Women					
Age at baseline assessment (n = 273,405)					
– Under 45	22,048 (10.3)	157 (22.4)	1130 (21.9)	278 (12.4)	4222 (8.4)
– 45–49	29,263 (13.6)	181 (25.9)	1224 (23.7)	344 (15.3)	5829 (11.6)
– 50–54	34,424 (16.0)	124 (17.7)	1064 (20.6)	359 (16.0)	7300 (14.5)
– 55–59	40,495 (18.8)	94 (13.4)	849 (16.4)	379 (16.8)	8939 (17.7)
– 60–64	52,227 (24.3)	91 (13.0)	652 (12.6)	498 (22.1)	12,637 (25.1)
– 65 and over	36,395 (16.9)	53 (7.6)	243 (4.7)	392 (17.4)	11,514 (22.8)
Deprivation (n = 273,078)					
– Least deprived	83,416 (38.9)	185 (26.5)	1097 (21.3)	535 (23.8)	15,773 (31.3)
– 2	45,987 (21.4)	141 (20.2)	824 (16.0)	431 (19.2)	9718 (19.3)
– 3	31,761 (14.8)	115 (16.5)	930 (18.0)	424 (18.8)	7514 (14.9)
– 4	27,607 (12.9)	107 (15.3)	1062 (20.6)	382 (17.0)	7757 (15.4)
– Most deprived	25,822 (12.0)	151 (21.6)	1241 (24.0)	478 (21.2)	9620 (19.1)
Lifetime number of sexual partners					
– Median (IQR)	3 (1–5)	3 (1–6)	10 (5–20)		
Men					
Age at baseline assessment (n = 229,138)					
– Under 45	18,672 (10.6)	456 (21.6)	712 (16.7)	255 (12.9)	3873 (8.8)
– 45–49	22,727 (12.9)	453 (21.4)	775 (18.1)	324 (16.4)	4960 (11.2)
– 50–54	25,898 (14.7)	356 (16.9)	753 (17.6)	347 (17.5)	5713 (12.9)
– 55–59	31,510 (17.8)	298 (14.1)	736 (17.2)	333 (16.8)	7190 (16.3)
– 60–64	42,853 (24.3)	310 (14.7)	774 (18.1)	420 (21.2)	11,025 (25.0)
– 65 and over	34,967 (19.8)	239 (11.3)	525 (12.3)	301 (15.2)	11,383 (25.8)
Deprivation (n = 228,842)					
– Least deprived	69,964 (39.7)	347 (16.5)	904 (21.2)	465 (23.6)	13,205 (29.9)
– 2	37,395 (21.2)	264 (12.5)	708 (16.6)	350 (17.7)	7889 (17.9)
– 3	25,494 (14.5)	321 (15.2)	581 (13.6)	355 (18.0)	6164 (14.0)
– 4	22,114 (12.5)	440 (20.9)	877 (20.5)	340 (17.2)	6691 (15.2)
– Most deprived	21,424 (12.1)	736 (34.9)	1203 (28.2)	463 (23.5)	10,148 (23.0)
Lifetime number of sexual partners					
Median (IQR)	4 (2–9)	15 (4–50)	10 (5–25)		

IQR: Interquartile range; n: Number.

sexual partners and MSMW a median of ten. UK Biobank participants with missing sexual history information were older, with a median age of 59 in women and 60 in men.

Table 4 shows the patterns of sexual behavior in UK Biobank compared with NATSAL-3. In general, the patterns were broadly comparable, with more women and men reporting same- and opposite sex sexual histories than exclusively same-sex behavior. However, the overall proportions of individuals with a history of ever having had same-sex sex were lower in UK Biobank, with 6.7% of women and 6.5% of men in NATSAL-3 compared with 2.7% of women and 3.5% of men in UK Biobank (Table 4), with the biggest difference among people reporting both same and opposite-sex sexual experiences.

In NATSAL-3, the proportions of individuals identifying as gay/lesbian, bisexual or other sexual orientations in were lower (2.0% in women and 2.6% in men) than the percentages of people who did not report an exclusively opposite-sex sexual history (Table 4). The PPV of exclusively having same-sex sex in detecting women and men who identify as gay or lesbian is high (68.3% in women and 82.6% in men). However, the PPV of a sexual history including both men and women for identifying people with bisexual sexual identity is low (9.1% in women and 13.7% in men). 15.6% of gay/lesbian women and 36.5% of gay/lesbian men are identified by considering people with an exclusively same-sex sexual history. The sensitivity of a history of sex with both men and women

Table 4. Sexual behavior and sexual identity in National Survey of Sexual Attitudes and Lifestyles and UK Biobank.

	Sexual behavior		Sexual identity	
	UK Biobank (%) (95% CI)	NATSAL-3 Weighted (%) (95% CI)		NATSAL-3 Weighted (%) (95% CI)
Women	**n = 220,714**	**n = 3260**		**n = 3266**
– Exclusively opposite-sex	97.3 (97.3–97.4)	93.3 (92.3–94.1)	Heterosexual/straight	98.0 (97.5–98.5)
– Exclusively same-sex	0.3 (0.29–0.34)	0.2 (0.1–0.5)	Gay/lesbian	1.0 (0.7–1.4)
– Both same- and opposite-sex	2.3 (2.3–2.4)	6.4 (5.6–7.4)	Bisexual	0.7 (0.4–1.1)
			Other	0.3 (0.2–0.6)
Men	**n = 183,014**	**n = 2329**		**n = 2357**
– Exclusively opposite-sex	96.5 (96.4–96.6)	93.5 (92.4–94.5)	Heterosexual/straight	97.4 (96.6–98.0)
– Exclusively same-sex	1.2 (1.1–1.2)	0.6 (0.4- 1.0)	Gay/lesbian	1.4 (1.0–1.9)
– Both same- and opposite-sex	2.3 (2.3–2.4)	5.8 (4.9–7.0)	Bisexual	0.9 (0.6–1.4)
			Other	0.3 (0.1–0.8)

n: Number; NATSAL-3: National Survey of Sexual Attitudes and Lifestyles.

Table 5. Positive predictive values and sensitivity of sexual behavior as a measure of sexual orientation – National Survey of Sexual Attitudes and Lifestyles.

Sexual behavior	Sexual orientation					
	Women			Men		
	Heterosexual/straight	Gay/lesbian	Bisexual	Heterosexual/straight	Gay/lesbian	Bisexual
Positive predictive value (95% CI)[†]						
– Exclusively opposite-sex	99.6 (99.2–99.8)			99.2 (98.6–99.6)		
– Exclusively same-sex		68.3 (33.0–90.4)			82.6 (47.1–96.2)	
– Both same- and opposite-sex			9.1 (5.8–14.0)			13.7 (9.0–20.3)
Sensitivity (95% CI)[‡]						
– Exclusively opposite-sex	94.6 (93.7–95.4)			94.7 (93.7–95.6)		
– Exclusively same-sex		15.6 (6.9–31.6)			36.5 (22.4–53.3)	
– Both same- and opposite-sex			85.6 (56.4–96.5)			84.3 (58.6–95.3)

[†]These positive predictive values present the % of women and men in each group defined by sexual behavior who identify with the most closely related sexual orientation (e.g., 68.3% of women who have only ever had sex with a woman identify as gay/lesbian).
[‡]These measures of sensitivity present the % of women and men who identify as each sexual orientation who report the most closely related sexual behavior history (e.g., 15.6% of women who identify as gay/lesbian report only ever having had sex with a women).

in identifying people with bisexual sexual orientation is 85.6% in women and 84.3% in men (Table 5). Analysis weights estimated from NATSAL-3 (the proportion of people with each sexual behavior history who report each sexual orientation), which can be used to correct analyses based on UK Biobank in order to estimate health inequalities experienced by sexual minority women and men are presented in Table 6, with a heterosexual/straight identity the most common for each sexual behavior (weights range from 0.749–0.997) except for an exclusively same-sex sexual history where a gay/lesbian identity has the highest weight.

Discussion

Main findings of this study

UK Biobank includes a large population-based sample of almost 3000 participants with a history of exclusively same-sex sexual behavior and over 9000 people with a history of sex with both men and women, with acceptable longitudinal consistency over repeated assessments. Reported history of same-sex sexual behavior overall was lower among UK Biobank participants than among respondents to NATSAL-3. Based on data from NATSAL-3, the

Table 6. Analysis weights (weighted proportions) estimated from National Survey of Sexual Attitudes and Lifestyles, which can be used to correct analyses based on UK Biobank in order to estimate health inequalities experienced by sexual minority women and men.

	Proportion from each group defined by their sexual behavior in the population who identifying as each sexual orientation			
	Heterosexual/straight	Gay/lesbian	Bisexual	Other
Sexual behavior measured in UK Biobank				
– Exclusively opposite-sex	0.997	0.000	0.001	0.002
– Exclusively same-sex	0.049	0.786	0.098	0.068
– Both same- and opposite-sex	0.749	0.130	0.113	0.008
– Never had sex	0.826	0.000	0.032	0.142
– Missing	0.846	0.154	0.000	0.000

appropriateness of using sexual behavior alone in UK Biobank (or other resources where sexual behavior, rather than sexual identity is measured) as a measure of sexual orientation is mixed; sexual behavior has high PPV and low sensitivity for sexual identity among women and men who exclusively have same-sex sex; a history of both same- and opposite-sex sex has low PPV and high sensitivity for bisexual sexual identity.

What is already known on this topic

Where previous research exists, our research findings are consistent. Our finding that WSWM have more sexual partners overall is consistent with previous analyses of a prior NATSAL survey [29]; as is our finding that people with a history of same-sex sexual behavior are younger, on average [24,30]. Our finding that WSW and MSM live in less affluent areas compared with WSEM and MSEW is consistent with previous research [1]. The link between material deprivation and poorer health is also well known [31], and so this finding additionally highlights the importance of addressing these health disparities. Although based on slightly different measures, our analyses of the intersection between sexual behavior and identity are also consistent with earlier work from the USA [23] and the UK [24].

Limitations of this study

Regarding the lower reported same-sex sexual history in UK Biobank compared with NATSAL-3, changes in sexual orientation and sexual behavior over time, with increased reporting of same-sex behavior and minority sexual orientations [22,30] are one possible explanation. A second is that people from less deprived areas are over-represented among UK Biobank respondents [17], while estimates from NATSAL-3 are nationally representative, and WSW and MSM on average live in more deprived areas [1]. A third explanation might be differences in the question wording across the two surveys, with a potentially broader definition of sexual experience in the NATSAL-3 survey question used in this analysis. Nonetheless the patterns of sexual behavior reported are consistent across the two datasets, and with prior external comparisons [18] and we argue that this is a reasonable validation of the UK Biobank responses. The availability of an independent data source (NATSAL-3) is a real strength for this work exploring the validity of UK Biobank as a tool to study sexual minority health disparities. A limitation of this study is that UK Biobank contains no self-reported measure of gender identity, although gender identity disorders in 90 cohort participants are captured through health record linkages [32]. We also note that many people have a sexual identity that is something other than heterosexual, gay/lesbian or bisexual [33] and finally that sexual orientation may represent a continuum rather than a set of mutually exclusive terms [34].

What this study adds: can we use UK Biobank as a research resource for studying sexual minority health?

The issues of the representativeness of UK Biobank for the whole UK population are well known [17], however, the value of the resource is also well established. Over 15,000 people report a history of ever having sex with someone of the same sex and over 12,000 participants can be identified as exclusively having same-sex sex or having sex with both women and men, on the basis of their responses to questions about lifetime numbers of sexual partners. This makes UK Biobank an incredibly important resource for sexual minority health research, as the depth of health, risk factor and other information is unique. The fact that respondents were not sampled on the basis of their sexual orientation and that there is a comparable group of participants who have never had same-sex sex is a second strength.

However, there are limitations to an approach that considers researching health disparities on the basis of sexual behavior rather than sexual orientation. First, this ignores the fact that in the UK it is sexual orientation rather than sexual behavior that is the protected characteristic under the 2010 Equality Act (and for which, there is therefore a statutory duty for public bodies to identify and address any inequalities) [11]. Second, there is the more fundamental challenge that research to address health inequalities should be person-centered, and focus on the identity rather than the behavior of the groups under study.

A second important limitation is that only approximately 10% of the participants identified in UK Biobank with a history of sex with both women and men will identify with a bisexual sexual orientation. Bisexual women and men have particularly poor mental health [1], however, this is a group that cannot be well identified in UK Biobank through the measure of sexual behavior available.

We cautiously suggest that UK Biobank is a valuable research resource for sexual minority health, although with limitations. We recommend that future waves of data collection with the UK Biobank cohort consider asking new questions about sexual orientation and gender identity to address in part these data gaps, but for researchers using data collected at baseline we present here three possible approaches for how best to use this resource.

Consider research based on sexual behavior rather than sexual orientation

Although the causes of health disparities experienced by sexual minority women and men are complex, for the long-term health conditions that UK Biobank is designed to study, it is plausible that some are associated with sexual behavior rather than sexual orientation, for example, the higher risks of human papillomavirus-associated oropharyngeal and anal and penile cancers [35]. Furthermore, although bisexual women and men are not well identified in UK Biobank, among women history of sex with both women and men is none the less a marker of adverse health outcomes, and this may well be an important group to study [29].

Consider only WSEW & MSEM

Although bisexual women and men are not well identified in UK Biobank, women and men who exclusively have same-sex sex has much higher PPV for gay/lesbian sexual orientation. For some research projects, this may be a valuable approach to take.

Use weights

The weights that we present in Table 6 provide a third possible approach for researchers. By creating new variables for each sexual orientation with these values based on sexual history, and using these new variables as the predictors in a model, adjusted estimates of disparities by sexual orientation can be estimated [36,37].

Conclusion

Given the pressing need for research to understand how to address disparities in health and health outcomes experienced by lesbian, gay and bisexual women and men, UK Biobank provides an important and to date underutilized resource for research. As the length of follow-up increases with time, the value will increase. However, there are limitations, particularly the lack of a measure of sexual orientation. Research based on sexual behavior only, limiting research to people with exclusive same-sex sexual history or research using the weights provided in this paper present three possible solutions to this issue.

Acknowledgments

The authors thank J Usher-Smith for very helpful comments on an earlier draft of this manuscript.

Financial & competing interests disclosure

CL Saunders received salary support from the National Institute for Health Research (NIHR) School for Primary Care Research (project reference SPCR-2014-10043 Capacity Building Award 10a SPCR Postdoctoral Fellowship Award). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. The data used in this study were accessed through UK Biobank (application number 42861).

No writing assistance was utilized in the production of this manuscript.

Summary points

- Despite poorer health and healthcare outcomes experienced by lesbian, gay and bisexual women and men, data for research to characterize and address these disparities remain limited.
- UK Biobank is a cohort study of over half a million participants, designed as a resource to study chronic and long-term health conditions; this analysis evaluates its strengths and limitations for sexual minority health research.
- We describe same- and opposite-sex sexual history information collected from 502,543 UK Biobank participants recruited between 2006 and 2010, and compare this with responses to the third National Survey of Sexual Attitudes and Lifestyles (NATSAL-3), collected between 2010 and 2012.
- After exclusions, 700 (0.3%) women and 2112 (1.2%) men in UK Biobank reported a sexual history of exclusively same-sex sex and 5162 (2.3%) women and 4275 (2.3%) men reported a sexual history of sex with both women and men.
- Estimates history of same-sex sex were slightly lower in UK Biobank than patterns seen in NATSAL-3, but longitudinal consistency in sexual history reporting was high (>98%).
- Analysis of NATSAL-3 found the positive predictive value of exclusive same-sex history for gay/lesbian sexual orientation was 68.3% in women and 82.6% in men, a history of sex with both men and women has positive predictive value of 9.1% in women and 13.7% in men for bisexual sexual orientation.
- UK Biobank is an important resource, however, there are limitations, particularly the lack of a measure of sexual identity.
- Research based on sexual behavior only, limiting research to people with exclusive same-sex sexual history or using the weights estimated here present three possible approaches to maximizing the use of UK Biobank as a resource for sexual minority health research.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Elliott MN, Kanouse DE, Burkhart Q *et al.* Sexual minorities in England have poorer health and worse health care experiences: a national survey. *J. Gen. Intern. Med.* 30, 9–16 (2015).
2. Meads C, Martin A, Grierson J, Varney J. Systematic review and meta-analysis of diabetes mellitus, cardiovascular and respiratory condition epidemiology in sexual minority women. *BMJ Open* 8, e020776 (2018).
3. Pakianathan M, Daley N, Hegazi A. Gay, bisexual, and other men who have sex with men: time to end the fixation with HIV. *BMJ* 354, i4739 (2016).
4. MacMillan Cancer Support. LGBT people with cancer: the emerging picture (2016). http://be.macmillan.org.uk/Downloads/CancerInformation/RichPicture/RP-LGBT-people-with-cancer.pdf
5. Institute of Medicine. The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding. (2011). www.ncbi.nlm.nih.gov/books/NBK64806/
6. Government Equalities Office. LGBT Action Plan (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721367/GEO-LGBT-Action-Plan.pdf
7. Cahill SR, Makadon HJ. If They Don't Count Us, We Don't Count: Trump Administration rolls back sexual orientation and gender identity data collection. *LGBT Health* 4, 171–173 (2017).
8. Benchimol EI, Smeeth L, Guttmann A *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 12, e1001885 (2015).
9. Brooks H, Llewellyn CD, Nadarzynski T *et al.* Sexual orientation disclosure in health care: a systematic review. *Br. J. Gen. Pract.* 68, e187–e196 (2018).
10. Cathcart-Rake EJ, Zemla T, Jatoi A *et al.* Acquisition of sexual orientation and gender identity data among NCI Community Oncology Research Program practice groups. *Cancer* 125, 1313–1318 (2019).
11. HM Government. Equality act (2010). www.legislation.gov.uk/ukpga/2010/15/pdfs/ukpga_20100015_en.pdf
12. Semlyen J, Hagger-Johnson G. Sampling frame for sexual minorities in public health research. *J. Public Health (Oxf.)* 39, 644 (2017).
13. Semlyen J. Recording sexual orientation in the UK: pooling data for statistical power. *Am. J. Public Health* 107, 1215–1217 (2017).
14. Government Equalities Office. National LGBT Survey: research report (2018). www.gov.uk/government/consultations/national-lgbt-survey

15. Hickson F, Davey C, Reid D, Weatherburn P, Bourne A. Mental health inequalities among gay and bisexual men in England, Scotland and Wales: a large community-based cross-sectional survey. *J. Public Health (Oxf.)* 39, 266–273 (2017).
16. Allen N, Sudlow C, Downey P *et al.* UK Biobank: current status and what it means for epidemiology. *Health Policy Techn.* 1, 123–126 (2012).
17. Fry A, Littlejohns TJ, Sudlow C *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* 186, 1026–1034 (2017).

• **Describes the characteristics of UK Biobank participants in detail.**

18. Ganna A, Verweij KJH, Day FR *et al.* Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science* 365, aat7693 (2019).
19. Perez-Cornago A, Key TJ, Allen NE *et al.* Prospective investigation of risk factors for prostate cancer in the UK Biobank cohort study. *Br. J. Cancer* 117, 1562–1571 (2017).
20. Saunders CL, Massou E, Waller J, Meads C, Marlow L, Usher-Smith JA. Cervical screening attendance and cervical cancer risk among women who have sex with women. *J. Med. Screen.* doi:10.1177/0969141320987271 (2021) (Epub ahead of print).

•• **An applied paper using UK Biobank looking at screening uptake among women who have sex with women.**

21. Erens B, Phelps A, Clifton S *et al.* Methodology of the third British National Survey of Sexual Attitudes and Lifestyles (NATSAL-3). *Sex. Transm. Infect.* 90, 84–89 (2014).
22. ONS. Sexual orientation, UK: 2019 (2021). www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/sexuality/bulletins/sexualidentityuk/2019
23. Mishel E. Intersections between sexual identity, sexual attraction, and sexual behavior among a nationally representative sample of American men and women. *J. Official Stat.* 4, 859–884 (2019).
24. Geary RS, Tanton C, Erens B *et al.* Sexual identity, attraction and behaviour in Britain: the implications of using different dimensions of sexual orientation to estimate the size of sexual minority populations and inform public health interventions. *PLoS ONE* 13, e0189607 (2018).
25. University College London, Centre for Sexual Health and HIV Research, Johnson, A. National Survey of Sexual Attitudes and Lifestyles, 2010-2012, 2nd Edition. (2018). [Data Collection]. UK Data Service. SN: 7799, http://doi.org/10.5255/UKDA-SN-7799-2

• **Research data available from here.**

26. National Centre for Research Methods. Townsend deprivation index (2021). www.restore.ac.uk/geo-refer/36229dtuks00y19810000.php
27. UK Biobank. Data-Field 189. Townsend deprivation index at recruitment (2021). https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=189
28. Crowe S, Barker E, Roberts M *et al.* Preprint: are we asking the right questions? Working with the LGBTQ+ community to prioritise healthcare research. *Research Square* DOI: 10.21203/rs.3.rs-501972/v1 (2021).
29. Mercer CH, Bailey JV, Johnson AM *et al.* Women who report having sex with women: British National probability data on prevalence, sexual behaviors, and health outcomes. *Am. J. Public Health* 97, 1126–1133 (2007).
30. Mercer CH, Tanton C, Prah P *et al.* Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (NATSAL). *Lancet* 382, 1781–1794 (2013).

• **Describes the characteristics of National Survey of Sexual Attitudes and Lifestyles participants in detail.**

31. Marmot M, Allen J, Boyce T, Goldblatt P, Morrison J. *Health equity in England: the Marmot Review 10 years on.* Institute of Health Equity, London, UK (2020).
32. UK Biobank. Data-Field 130940. Gender identity disorders (2021). http://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=130940
33. Elliott MN, Dahlhamer JM, MacCarthy S *et al.* Using ancillary sociodemographic data to identify sexual minority adults among those responding "something else" or "don't know" to sexual orientation questions. *Med. Care* 57, e87–e95 (2019).
34. ONS. Measuring sexual identity: an evaluation report (2011). https://data.gov.uk/dataset/e6033627-3106-47a9-a3fb-4bcaa34c1bcb/measuring-sexual-identity-evaluation-report
35. Saunders CL, Meads C, Abel GA, Lyratzopoulos G. Associations between sexual orientation and overall and site-specific diagnosis of cancer: evidence from two National Patient Surveys in England. *J. Clin. Oncol.* 35, 3654–3661 (2017).
36. Elliott MN, Fremont A, Morrison PA, Pantoja P, Lurie N. A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity. *Health Serv. Res.* 43, 1722–1736 (2008).
37. McCaffrey DF, Elliott MN. Power of tests for a dichotomous independent variable measured with error. *Health Serv. Res.* 43, 1085–1101 (2008).